

Article

International immigration distribution at local authority level: research into potential improvements

Research into whether accounting for lags in international immigrants appearing on the NHS Patient Register can improve the mid-year population estimates.

Contact:
Neil Park
pop.info@ons.gov.uk
+44 (0)1329 444661

Release date:
27 July 2020

Next release:
To be announced

Table of contents

1. [Main points](#)
2. [Summary](#)
3. [Distributing International immigration flows to local authority level](#)
4. [Research into administrative data lagging](#)
5. [Analysing the impact of migrant lags on local authority distributions](#)
6. [Identifying other migrants who may appear in previous years as working or student immigrants](#)
7. [Impact on local authority distribution of migration](#)
8. [Glossary](#)
9. [Limitations of linkage](#)
10. [Limitations of administrative data](#)
11. [Conclusions](#)

1 . Main points

- The method currently used to distribute international immigrant flows to local authorities cannot be improved by accounting for new insights into lagging in administrative data for migrants.
- Around 10,000 NHS Patient Register records used to distribute the flow of non-British 17- to 59-year-old, non-working, non-student immigrants (referred to as “other migrants”) in mid-2018 could be found as workers or students in previous years.
- Removing these records from the distribution of other migrants has no significant impact on the resulting international immigration estimates at local authority level.
- No changes will be made to the method used to distribute international immigrants to local authorities.

2 . Summary

The mid-year population estimates for England and Wales use a mixture of survey, administrative and census data to distribute national international immigration flows to local authority level. For all [processing since 2011](#), the principle has been to take the national immigration flows (long-term international migration), broken down by reason for migration, and use the most appropriate administrative data to distribute the flow to local authorities. The three main administrative data sources used in this process are:

- the [Migrant Worker Scan \(MWS\)](#) (for working migrants)
- data on higher education from the [Higher Education Statistics Agency \(HESA\)](#)
- the GP [NHS Patient Register](#) (for other migrants)

These three data sources are available to the Office for National Statistics (ONS) at a record level, and work is carried out to link them together so that duplication across the data used to distribute workers, students and other migrants can be addressed.

In January 2019, we published [research](#) into new international immigrants’ patterns of registration for a National Insurance number (NINo) and the NHS. Analysis showed new international immigrants do not register themselves for a NINo and with doctors surgeries at the same time. The lag between registrations could mean they appear in different reporting periods on different datasets. Given our use of linked administrative data sources, this means we could potentially include the same migration event, via different administrative data sources, in two successive years of our immigration distributions.

We have conducted research to assess whether lags in registering for administrative data sources impact our existing method for distributing international immigrants to local authorities. This research showed that we could link 10,000 (out of 180,000) of the GP patient registrations with a Flag 4 to MWS records from previous years. However, removing these records has a negligible impact on the resulting estimates of local authority international immigration as they are spread relatively evenly across England and Wales.

3 . Distributing International immigration flows to local authority level

Our current method for producing international immigration flows at local authority level involves taking the immigration flow for England and Wales, by reason for migration, and distributing this to local authorities. This is completed using the most appropriate administrative data available. Where possible, we conduct record-level linkage between administrative data sources. This ensures that each migrant in a given year only features in the distribution for a single reason for migration as each new migrant can appear in one, two or all three of the main datasets used to distribute the immigration flows. By ensuring migrants only appear on one dataset for distribution purposes, we reduce the potential for bias in our distributions.

In practice, this means we link together the Migrant Worker Scan (MWS), data from the Higher Education Statistics Agency (HESA) and data from GP registrations of Flag 4s to new migrants. Additional information on the methods used to produce immigration estimates at local authority level are provided in the [mid-year population estimates methodology guide](#).

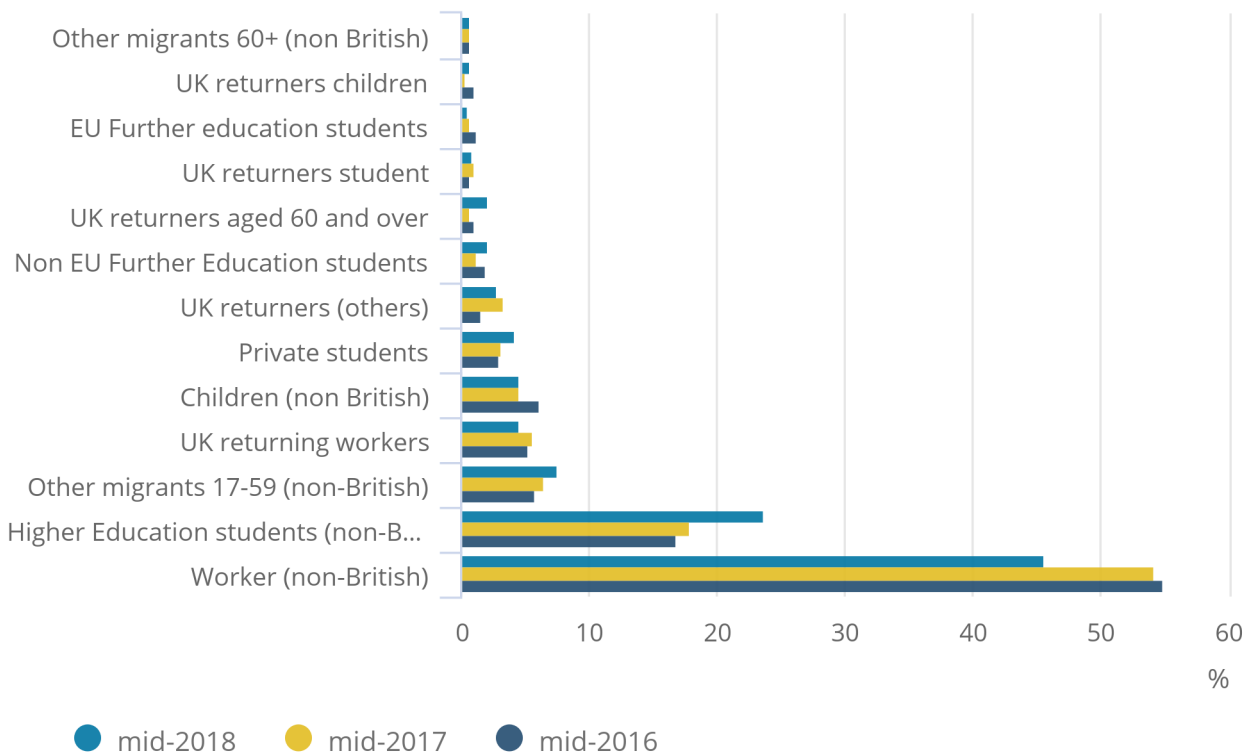
Figure 1 shows the highest proportion of immigrants each year come to England and Wales for work followed by higher education study. Those migrants aged 17 to 59 years coming to England and Wales for other reasons, and distributed to local authority using the NHS Patient Register, account for around 7% of the total flow. Around 20% of the immigrant flow (after excluding new asylum seekers and refugees) is made up of migrant children, those aged over 60 years, students at further education colleges or private institutions, and/or returning UK migrants.

Figure 1: Other migrants aged 17 to 59 years are the third largest immigration stream

International immigration streams by proportion of overall flow, England and Wales, mid-2016 to mid-2018

Figure 1: Other migrants aged 17 to 59 years are the third largest immigration stream

International immigration streams by proportion of overall flow, England and Wales, mid-2016 to mid-2018



Source: Office for National Statistics

Notes:

- 1. Immigration flows exclude asylum seekers and refugees are not shown.

4 . Research into administrative data lagging

The [research we published in January 2019](#) showed clear differences between when new international immigrants appeared on administrative data sources. These included:

- EU nationals register more quickly for a National Insurance number (NINo) than non-EU nationals
- non-EU nationals register more quickly with the NHS than EU nationals (EU median lags: 276 days; non-EU median lags: 60 days)
- females tend to register more quickly with the NHS than males, with this trend seen in all except the youngest age group (males' median lags: 291 days; females' median lags: 150 days)

These findings imply a strong likelihood a proportion of migrants appearing as Flag 4s in one year may appear as worker or student migrants in other years. In effect, the same migration event could be informing multiple years of our immigration distribution, potentially leading to bias in our estimates.

However, because of the way our processing accounts for immigrants coming for work and study, it is highly unlikely that any migration events for these would appear in the other migrant stream the year before. This is because working migrants are only included in our distributions if their arrival date in the UK was within the reference period. Consequently, if a migrant did appear on the previous year's NHS Patient Register as a Flag 4 and subsequently applied for a NINo in this year's reference period, they would be excluded from our distribution of workers because of their arrival date. For Flag 4s, we only receive the date they registered with a GP rather than a self-reported arrival date.

5 . Analysing the impact of migrant lags on local authority distributions

The findings of the [January 2019 research](#) show a clear need to assess the potential impact of this issue on the current production of population estimates. The immigration distribution system produces annual datasets that hold the records from each data source that are likely to have been new international immigrants in the reference period. This means that for each year, we have access to three files:

- a “clean” worker migrant file based on the Migrant Worker Scan (MWS) with higher education students removed
- a “clean” higher education student file based on Higher Education Statistics Agency (HESA) data with workers removed
- a “clean” other migrant file based on the Flag 4s from the NHS Patient Register with workers and students removed

Identifying lagged migrant registrations

The approach taken was to link the other migrants in the most recent year (mid-2018) with the workers and students from the previous year (mid-2017) using the linking methodology already in place. This methodology uses a set of match keys to allow anonymised administrative data sources to be linked together. Several match keys are used, ranging from an exact match of full name, date of birth, sex and postcode through to less stringent match keys allowing for changes and minor errors in each variable. The full list of match keys used is shown in Table 1.

When datasets are linked together, diagnostic information on the match keys used for linkage is produced. This information showed that in the majority of cases, links between years were made using the most stringent match keys. This provides confidence that the matches between years were genuine (the same person in two different years) rather than false matches.

Table 1: Description of match keys used

Match key	Details
1	Forename, Surname, DOB, Gender, Postcode
12	Forename, Surname, DOB, Gender, Postcode district
13	Forename, Surname, DOB, Gender, Postcode area
17	Forename, date of birth, gender, postcode
3	Forename Initial, Surname Initial, DOB, Sex, Postcode District
7	Forename bi-gram, Surname bi-gram, DOB, Sex, Postcode Area
4	Forename Initial, DOB, Sex, Postcode
5	Surname Initial, DOB, Sex, Postcode
11	Forename, Surname, DOB, Postcode
6	Forename, Surname, Sex, Postcode
2	Forename, Surname, DOB, Sex
8	Forename, Surname, Year of Birth, Sex, Postcode District
9	First Middle Name, Surname, DOB, Sex, Postcode
10	Second Middle Name, Surname, DOB, Sex, Postcode

Source: Office for National Statistics

Assessing the potential for longer lags

Following this, we linked the residual other migrants from mid-2018, after removing the mid-2017 linked records, with the workers and students for mid-2016 and then repeated the process for mid-2015. Linking to mid-2016 and mid-2015 allowed the possibility of very long lags to be assessed.

The final linkage involved taking the other migrants from mid-2017 and linking these to the workers and students from mid-2017 to check that patterns found in the most recent year were replicated in other years.

6 . Identifying other migrants who may appear in previous years as working or student immigrants

In carrying out the linkage work, we verified that international immigrants can appear in different administrative data sources in successive years and that the same migration event can influence two years of our immigration distributions. However, this work also shows that removing these records from the most recent year's migration data has a negligible impact on the resulting distribution of international migrants that we use in the construction of population estimates.

Table 2 shows that around 10,000 of the Flag 4s used to distribute other migrants for mid-2018 linked to either a worker or student record for either mid-2017 (T minus 1), mid-2016 (T minus 2) or mid-2015 (T minus 3). At a high level, this research confirmed the findings of the [January 2019 research](#) with the majority of those having relatively short lags between their appearance on different administrative data sources. In our linkage of mid-2018 Flag 4s with previous years' data, we found most links between mid-2018 and mid-2017 (7,000 links) and progressively fewer links to each earlier year's data (2000 linked to mid-2016 and 1,000 to mid-2015).

To check how representative the findings were for other years, the first stage of the analysis (linking to T minus 1) was repeated using mid-2017 as the base year (see Table 2). The linkage of mid-2017 Flag 4s (used to distribute other migrants) to mid-2016 gave confidence that the patterns observed in mid-2018 were a good representation of the scale of this issue for other years.

Table 2: Linkage of other migrant records to previous years' workers and students

	Mid-2018 Other migrants	Mid-2017 Other migrants
Total other migrant records	180,000	180,000
Linked to T-1	7,000	7,000
Linked to T-2	2,000	N/A
Linked to T-3	1,000	N/A

Source: Office for National Statistics

Notes

1. Linkage of mid-2017 other migrants to T minus 2 and T minus 3 was not carried out. [Back to table](#)
2. T minus 1 refers to one year before the reference year. For mid-2018, this means mid-2017. For mid-2017, this means mid-2016. [Back to table](#)

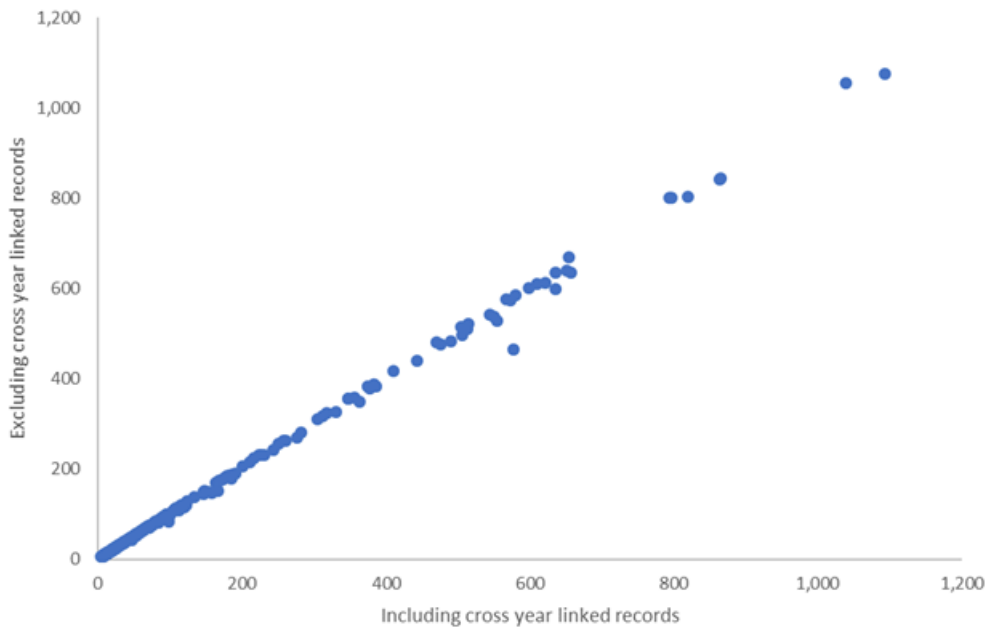
7 . Impact on local authority distribution of migration

As discussed in [Section 6: Identifying other migrants who may appear in previous years as working or student immigrants](#), around 6% (10,000) of Flag 4s used to distribute other migrants in mid-2018 could be linked to a Migrant Worker Scan (MWS) or Higher Education Statistics Agency (HESA) record from mid-2017, mid-2016 or mid-2015. The next stage of the project was to assess the impact of removing these records from our production of local authority immigration and population estimates. As part of the linkage process, records that could be identified in previous years' data were flagged to allow them to be removed from the distribution of other migrants. Following the removal of these lagged administrative records, we recalculated the local authority migration distribution.

As Figure 2 shows, there was a very strong relationship between the distributions of other migrants both before and after the removal of lagged administrative records. This demonstrates that the removal of lagged administrative records from the distribution of other migrants did not substantially alter the distribution of other migrants.

Figure 2: Excluding cross-year linked records only has a small impact on other migrant flows

Distribution of other migrants aged 17 to 59 years including and excluding cross-year linked migrant records, England and Wales, mid-2018

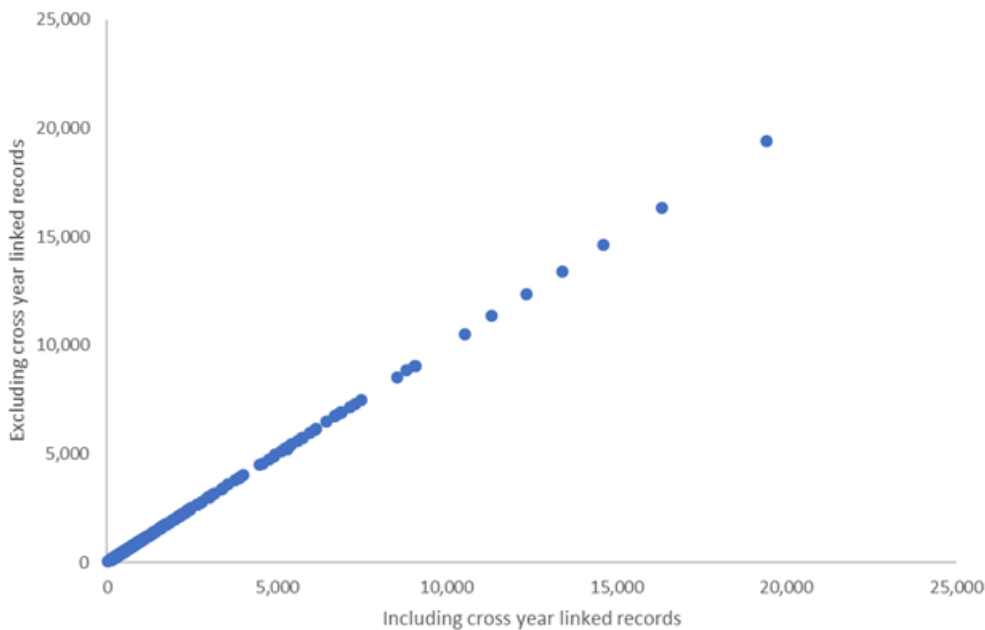


Source: Office for National Statistics

Once all streams of international immigration are incorporated together, the impact of removing linked records from the other stream is negligible. Were these new figures to be incorporated, the impact on immigration flows would be less than positive or negative 1% (plus or minus 40 people) in all but one local authority. The one exception is Cardiff where the immigration flows would be reduced by around 2.1% (100 people).

Figure 3: Local authority immigration flows are largely unchanged by removing cross-year linked records

Distribution of international immigrants including and excluding cross-year linked migrant records, England and Wales, mid-2018



Source: Office for National Statistics

Notes:

1. The data in this chart shows immigration flows excluding the inflow of asylum seekers and refugees.

It is important to note that the potential impact of issues with the international immigration distribution can compound over the 10 years of the intercensal period and that small differences in a single year can become more significant after 10 years. However, even in Cardiff, the total impact on the population estimates after 10 years is still likely to be small. In our research, we looked at both the impact on mid-2018 and mid-2017 immigration flows and found that the impacts of removing linked records were broadly similar. Based on these findings, we can reasonably assume that the impact is likely to be similar in scale each year; one plausible scenario would be that the impact of the issue over the whole decade may be in the region of 1,100 for Cardiff, the most affected local authority. Given the population of Cardiff in mid-2018 was 364,000, this would make a difference of around 0.3%. For reference, the latest uncertainty measures for Cardiff's population estimates (for mid-2016) were plus or minus 3%.

8 . Glossary

Migrant Worker Scan (MWS)

The Migrant Worker Scan (MWS) contains information on all adult overseas nationals who have registered for and been allocated a National Insurance number (NINo). A NINo is generally required by any overseas national looking to work or claim benefits or tax credits in the UK.

Higher Education Statistics Agency (HESA) data

Higher Education Statistics Agency (HESA) data provide the only comprehensive nationally consistent source of data on higher education students. HESA data provide basic demographic data (age and sex); term-time and domicile (parental home) postcode; information on the course (course length and type); and institutional information.

NHS Patient Register

The NHS Patient Register is a record of all persons registered with a GP in England and Wales. The NHS Patient Register was used to maintain an accurate list of all persons registered with a GP, allowing the timely transfer of medical records and correct payments to doctors. It contains a list of everyone who is, or ever has been, registered with a GP in England and Wales since the NHS was founded in July 1948.

Flag 4

The NHS Patient Register holds a “flag” for persons whose previous place of residence was outside of the UK. These data are used (together with other sources) to help allocate separate estimates of international migration into the UK to local authority areas.

Match key

Data linkage is a process that temporarily brings together two or more sets of administrative or survey data from different organisations. Match keys are created by putting together pieces of information to create unique keys that can be hashed and used for automated data linkage, with the intention of eliminating some of the discrepancies that might otherwise prevent an automated match. For example, a match key might be constructed from the first three characters of an individual’s forename and surname, combined with their date of birth, sex and postcode district.

Lag

The time difference between an event happening (such as an individual migrating between countries) and that event being reflected in administrative data sources is referred to as the lag.

9 . Limitations of linkage

The linkage work involves [matching anonymised administrative records using match keys \(PDF, 319KB\)](#). This has clear limitations as it is likely that some false matches will be made and some genuine matches will be missed.

Further, there is the potential for linked administrative records to relate to different migration events of the same person. For example, an individual could migrate to England and Wales in June 2016, emigrate in June 2017 and return in June 2018.

10 . Limitations of administrative data

As shown in [Section 7: Impact on local authority distribution of migration](#), the impact of removing lagged records on the local authority distribution of migrants is small. The most obvious inference from this is that migrants behave in similar ways irrespective of where in England and Wales they migrate to. However, it is worth considering that the impact could be larger if the administrative data more precisely met our target definitions and concepts. One issue is that the administrative data include records of several groups that fall outside of our criteria of long-term international migrants. It is likely that this explains why the 40,000 other non-UK migrants are distributed to local authorities in England and Wales using 180,000 Flag 4s. Some of the reasons for the discrepancy between Flag 4s and the other migrant flow include:

- short-term migrants (those who remain in the country for less than 12 months) applying for GP services
- UK-born returning migrants being issued Flag 4s
- missed links between the Migrant Worker Scan (MWS), Higher Education Statistics Agency (HESA) data and the GP NHS Patient Register
- incorrectly issued Flag 4s (people moving within the UK who are issued with Flag 4s)

The excess records in each data source may be dampening down the impact of removing linked records.

11 . Conclusions

This work has shown that the method for distributing international immigrants to local authorities in the mid-year estimates has not been negatively impacted by different levels of lagging across administrative data sources.

Further, the immigration distribution as currently used in the production of the mid-year estimates cannot be meaningfully improved at the present time through the inclusion of cross-year linkage of administrative data. While some records are influencing the immigration distribution in multiple years, the removal of records does not result in any significant change in the amount of immigration at the local authority level.

Part of the explanation for the relatively minor impact of making this change is that the other migrant flow is relatively minor compared with the workers and students flows. Given the relatively minor impacts indicated by this research, no changes will be made to the processing of the mid-year estimates.

As part of the Population Statistics and Migration Transformation, further research into better ways of producing immigration and emigration flows are ongoing, and research will continue to be published.